The effect of financial rewards on students' achievement: Evidence from a randomized experiment[1]

Edwin Leuven Hessel Oosterbeek Bas van der Klaauw

[1]The authors are affiliated with research program 'Scholar' and the Tinbergen Institute. Leuven and Oosterbeek are at the Department of Economics of the University of Amsterdam. Van der Klaauw is at the Department of Economics of the Free University in Amsterdam. We thank the dean of the Department of Economics of the University of Amsterdam Jacques van der Gaag who gave us permission to conduct the experiment reported in this paper. We also thank seminar participants in Amsterdam, Sevilla, Stockholm and Tilburg for their comments.

Abstract

This paper reports about a randomized field experiment in which first year economics and business students at the University of Amsterdam could earn financial rewards for passing the first year requirements within one year. Participants were assigned to a high, low and zero (control) reward group. The passing rate and the numbers of collected credit point are not statistically different across the three groups. We do find some evidence for heterogeneous treatment effects. In particular, high ability students and students from higher social backgrounds have higher passing rates and collect more credit points when assigned to (higher) reward groups. Students in the reward groups, however, do not report to have studied more hours.

Keywords: financial incentives, student achievement, randomized social experiment, heterogeneous treatment effects, university education

JEL Codes: I21, I22, J24

1. Introduction

Universities in the Netherlands are public and the system of university education is characterized by low tuition fees. Undergraduate students collect credit points by passing exams, which are graded on a scale from 1 (lowest) to 10 (highest). To pass an exam a score of at least 6 is required. When students fail an exam (score below 6), they do not collect any credits until they pass a make-up exam. Failing an exam is common, and many students do more than one make-up exam for the same course before passing. As a consequence, most students do not graduate within the nominal duration of the program (4 years). Delay typically starts in the first year. Moreover, for teachers the system of failing and passing exams implies much grading since they often grade multiple exams of a student for the same course.

The share of undergraduate students in economics and business at the University of Amsterdam who pass all first year requirements within their first academic year is in the vicinity of 0.20 (ranging between 0.17 and 0.22 in the last five years). For the Department of Economics and Business this low passing rate is a concern since public funding depends among other things on the number of credits points awarded each year. Moreover, once a year a Dutch weekly magazine (Elsevier) publishes a ranking of university departments in each field aimed at secondary education students who are in the process of choosing their university education. The first year passing rate is one of the inputs of this ranking.

During the past years the department has tried to improve the first year passing rate by introducing extra guidance by older students, increased teaching hours and more intensive courses. Passing rates have, however, not increased. The idea is that without additional effort from students passing rates will remain low, and that to improve the passing rates policy should aim at increasing students' effort. In this paper we study financial incentives as an instrument to achieve this.

A clear indication that financial incentives may have substantial effects on achievement comes from a study held in the academic year 1999/2000 at the University of Amsterdam. At the beginning of the third trimester, all first year students who followed the undergraduate program in econometrics were promised a reward of €454 (1000 Dutch guilders) upon fulfilling all first year requirements before the start of the new academic year.[1] In the year that this reward was in place, the passing rate was 0.50, while in the previous year this was 0.28 (cf. Hilkhuysen 2000). Those involved in the design and evaluation of this study attributed this increase of the passing rate to the reward. While the 0.22 increase in the passing rate may be the causal effect of the reward, this need not be the case. Plausible alternative explanations for the increased passing rate are a higher quality of the student cohort, less demanding courses, and less strict grading of exams. Given the design of the study it is difficult to establish a causal relation between the financial incentive and the increased passing rate. Nevertheless the results suggest that a financial incentive may be a very effective intervention.

In this paper we explore this further by reporting on a field experiment where first year undergraduate students in economics and business at the University of Amsterdam were randomly assigned to three groups. Students assigned to the 'high reward' group were promised a bonus of €681 on completion of all first year requirements by the start of the new academic year. Students assigned to the 'low reward' group were promised a bonus of €227 for this achievement. Students assigned to the control group could not earn a reward. Such a randomized experiment should provide a more convincing estimate of the causal effect of financial incentives on students' achievement than the related study among econometrics students discussed above.

To our knowledge, Angrist and Lavy (2002) is the only other study to date that analyzes the effects of financial rewards on students' achievement in an experimental setting.[2] They evaluate the effectiveness of financial incentives on obtaining secondary education matriculation in Israel. They implemented two experiments targeted at low-achieving students. Their first (pilot) experiment randomized 489 students within schools, of which 248 students were assigned to the treatment group. Treated students could earn a reward of $800 in cash (or $1000 to $1200 in education vouchers) upon completing their secondary education matriculation certificate. To obtain the support of school administrators, the randomization favored potentially low-achieving students toward the treatment group. Using instrumental variables estimation, Angrist and Lavy do not find a significant effect of the reward on achievement.

Their second (follow-up) experiment is a school-based experiment in which 20 out of 40 entire schools (with low matriculation rates) are assigned to treatment. According to the original design, students in treated schools could earn up to $2500 in cash during a three-year period. However, the program was abolished after one year. As a result students in grades 10 and 11 could earn at most $500 and students in grade 12 at most $1500. Angrist and Lavy (2002) find that treated schools have matriculation rates of 6-8 percentage points higher than untreated schools.[3]

A feature that the experiment in this paper has in common with the first experiment of Angrist and Lavy (2002) is that randomization takes place within the educational institution. Reward sizes are also of similar magnitude. Unlike Angrist and Lavy, who faced reluctant school administrators, we were in the position to implement a genuine randomized assignment. Moreover, we collected information about students' study effort. This permits us to also examine the impact of the rewards on effort.

To briefly summarize our results, for the full sample, we find no effect of the rewards on achievement measured by passing rates and numbers of collected credit point. Our results suggest that the absence of an effect on achievement can be attributed to students not spending more time on their study. While the average treatment effect is approximately zero, we find some evidence for heterogeneous treatment effects. In particular, students with high math skills and students with higher educated fathers have higher passing rates and collect more credit points when assigned to (higher) reward groups. While reported study time for these groups is not affected by treatment status, these students claim that they have studied harder as a consequence of the rewards. The positive effect for students with good math skills suggests that the average treatment effect is negligible not because the average student is unresponsive to financial incentives but because the requirements for the rewards are too demanding for the average student when compared to the size of the (uncertain) reward.

The remainder of this paper is organized as follows. The next section provides relevant background information about the Dutch system of higher education and of the economics and business program at the University of Amsterdam. Section 3 explains the design of the field experiment and describes the data. Section 4 presents and discusses the results, and Section 5 concludes.

2. Background

2.1 The Dutch system of university education

University education in the Netherlands is accessible for all students with a qualification from the pre-university track in secondary education.[4] This secondary education qualification can only be obtained by passing a uniform nationwide exam. The relevant secondary education exit requirements are set such that they are considered to be sufficient university entry requirements, and therefore all students starting a university education in economics or business are supposed to be capable of actually graduating (given that they exert sufficient effort).

In the academic year 2001/2002 there were 34,200 first year students at Dutch universities, which is about 17 percent of the relevant birth cohort. Some university studies (such as economics, history or mathematics) may require specific courses to be included in the secondary education curriculum. Apart from this, universities are not permitted to select students; everyone who applies with a valid entry qualification has to be admitted.[5] In the Netherlands selection therefore takes place at the exit of secondary education as opposed to at the entry of higher education.

Currently six Dutch universities offer an undergraduate program in economics and business.[6] While the programs offered by the different universities differ somewhat, they are considered to be close substitutes. They attract students from the same pool of secondary school graduates and they prepare their students for the same labor market, although people tend to stay in their region of origin. Oosterbeek et al. (1992) compare the labor market outcomes of graduates from the different economics and business departments in the Netherlands and find that selection corrected wage differentials are modest.

University students in the Netherlands who are younger than 30 and subscribed as full-time students are all charged the same tuition fee of €1,329.57 per year (in the academic year 2001/2002). The tuition fee is set by the government and does not vary by field of study or by university. There is also a uniform financial aid system that applies to all university students. The financial aid scheme consists of three components that students are entitled to for a maximum of four years. The first component is a basic grant of €211 per month for students who do not live with their parents and €69 per month for those students who do. The second component is an additional grant decreasing with parental income. The maximum size of this additional grant equals €222 per month. The third component is a loan. The maximum amount of this loan equals €456 per month.

An important feature of the basic grant and the additional grant is that they become loans if a student fails to collect enough credit points. Grants received during the first study year are not transformed into a loan if the student earns at least half of the nominal number of credit points of the first year or if the student obtains a higher education diploma within 10 years. The requirements for the grants not to turn into a loan are therefore not very demanding. In our population of economics students 58% of our population (see below) collects at least half of the credit points. Furthermore, of a given cohort of students over 80% will actually graduate within 10 years. Recall that the financial rewards in the experiment are only paid if a student collects all credit points, therefore the financial aid system of the government does not interfere with our financial rewards.

The complete loan plus interest (from the month of payment onward) must be repaid within 17 years after graduation. People with annual earnings below a certain threshold are exempted from repayment. Otherwise repayment is in fixed amounts independent of income. The loan component of the financial aid scheme is not very popular among Dutch students. Students typically use the basic grant and the additional grant, but of the total amount available for loans less than 20 percent is requested. This is reflected in the fact that many university students combine studying with some hours of paid work. In our sample around 80 percent of the students work, and they work on average around 12 hours per week (details concerning data collection are provided later).

2.2 Undergraduate program in economics and business at the University of Amsterdam

The undergraduate program in economics and business at the University of Amsterdam is a 4 year program. In the first academic year, which runs from September until August, all students in economics and business follow exactly the same program of 14 compulsory courses. The first year program was divided into three trimesters of 14 weeks each in the year that the experiment was conducted. Every trimester ended with exams shortly after the courses finished and the make-up exams are organized in the last week of August. The first academic year thus consisted of 42 study weeks, which are allotted to different courses in the form of credit points.[7] It is only after the first trimester of their second academic year that students choose different packages of courses to specialize either in economics or in business (there are also various sub-specializations within economics and within business)

Students typically surpass the nominal study duration of 4 academic years before obtaining their undergraduate degree. The delay of their study often starts already in the first year and passing rates are correspondingly low. For the Department of Economics and Business this is costly since the internal funding scheme of the University of Amsterdam is primarily based on the number of credit points that students obtain. Delay, which manifests itself in make-up exams and repeated attendance of the same courses, represents a net loss for the department because it increases the costs while no extra compensation is received.

3. Experimental design and data

As mentioned above, the experiment was motivated by the experience from an earlier study among econometrics students at the University of Amsterdam. When we proposed to run a randomized experiment, student representatives argued that experimentation with students is unethical and that the money involved with the experiment benefited only a small portion of all students of the department. Some of our colleagues supported this view. Nevertheless the dean of the department decided in favor of conducting the experiment under the requirement that participation ought to be voluntary. Voluntary participation was also necessary to obtain students' permission to access their records in the student administration.

In the experiment the rewards are tied to passing all exams before the second academic year starts. The first year passing rate is therefore the prime outcome variable of interest. We also consider the number of credit points collected by the students. Not only is it a relevant outcome measure since internal funding at the University of Amsterdam depends on this, but if the rewards give students an incentive to spend more effort on their study, we expect the number of credit points to increase also for students who do not pass all requirements. Note that since all courses in the first year are mandatory, students in the treatment groups cannot opt for less demanding courses.

The experimental design includes, beside the control group, a high reward group and a low reward group. The reward sizes of the high and low reward groups are €681 and €227, which is $1\frac{1}{2}$ and $\frac{1}{2}$ times the size of the €454 reward in the earlier study among econometrics students. Given the substantial increase in the passing rate attributed to the earlier €454 reward, the rewards in the present experiment seem sufficiently large to increase passing rates. At the same time, the size of these rewards are such that it would be feasible for the government or the university to implement the reward scheme if indeed the rewards result in higher passing rates and/or more realized credit points. The distinction in a high and low reward group allows us to distinguish between the effect of being treated as such, and the effect of the size of the reward. If both rewards induce the same effect, the amount of the reward is apparently not important within this range. If the high reward induces a larger effect than the low reward, it apparently matters how much can be earned.

To ensure that all students were treated identically, participation in the experiment was only open to students who (*i*) followed the full-time program, (*ii*) did not claim more than 1 credit point dispensation,[8] and (*iii*) did not start the economics and business program in a previous year. The total number of eligible students equals 254.

On October 1 2001, almost one month after classes started, we sent all first year students a letter inviting them to participate in the experiment. This was the earliest possible date given the availability of addresses from the student administration. The letter explained the purpose of the experiment and informed students that participants would be randomly assigned to three equally sized groups with equal odds for all students. Furthermore the letter explained that participation implied that the student granted the researchers permission to link information from the experiment to information from the student records about their achievements. Students received a fixed payment of €22.69 (50 Dutch guilders) upon participation. Notice that this procedure reveals clearly that no participant looses from the experiment. Everyone receives a small payment and everyone faces equal probabilities to be assigned to one of the reward groups. After a reminder and a telephone round 249 eligible students participated in the experiment, which is 98% of all eligible students. Three students could not be reached and 2 students explicitly rejected participation.

In the random assignment 83 students were assigned to the high reward group, 84 students to the low reward group, and 82 students to the control group. On November 29, letters were sent informing participants about their assignment status. The first exam was on November 28, the second on December 12 and the third on December 19.

Students had to fill out and sign a participation form which also included a short questionnaire. This questionnaire collected information about respondents' mathematics grades in secondary school and their parents' education.

Table 1 presents descriptive statistics of the background characteristics of the complete sample and for different subgroups. We distinguish two types of parental education, low and high, where high refers to all higher education, both university and vocational. Dutch pre-university secondary education offers two programs in mathematics: mathematics A and mathematics B. Mathematics A is considerably less advanced than mathematics B. Students are allowed to do exams in both programs, but it is not compulsory to do mathematics A in order to do mathematics B. Table 1 reports the shares of students who did exam for mathematics A only, for mathematics B only, and who did both mathematics A and B. Recall that exams are graded on a scale from 1 (lowest) to 10 (highest).

The random assignment was done by stratifying the participants on the basis of their mathematics results and their parents' education. This precludes that the random assignment procedure accidentally results in groups that differ in these observed characteristics.

The pre-assignment questionnaire also asked participants their subjective probability of fulfilling the requirement of passing all exams within the first academic year if they would be assigned to the control group, the low reward group and the high reward group respectively. This was done to get some indication of the effect of the rewards before the experiment actually took place. The average expected probabilities are reported in the bottom part of Table 1. Without a reward the expected passing rate equals 0.55. Given the actual passing rates from previous years of around 0.20, students seem overly optimistic at the beginning of their study. If students would be entitled to the low reward the expected passing rate increases to 0.63, and it increases to 0.71 for the high reward. This implies that ex-ante the students expected quite sizable effects from the rewards. No differences are observed across groups.

After the experiment ended a second questionnaire was sent to all participants. Upon completion, students received a payment of €25. In total 234 participants responded, which is 94% of all participants. This post-experiment questionnaire asked questions concerning students' current study status, the time they spent on their studies during the past year, their work activities during the past study year, their perceptions of the effect of the reward on their effort (if assigned to one of the reward groups), and possible supplementary rewards offered by third parties. We discuss the results below.

The sample size is not very large, which has implications for possible differences in passing rates between the groups that can be distinguished. To inves-

Table 1: Sample means of background characteristics (stratified by groups)

	All	High	Low	Control
Education father				
Higher education	0.53	0.54	0.52	0.54
Lower education	0.46	0.46	0.46	0.46
Education mother				
Higher education	0.37	0.36	0.39	0.37
Lower education	0.62	0.64	0.61	0.62
Only math A				
Share	0.60	0.61	0.60	0.59
Grade	6.7	6.7	6.7	6.8
Only math B				
Share	0.22	0.26	0.21	0.19
Grade	6.4	6.6	6.3	6.3
Both math A and math B				
Share	0.16	0.13	0.18	0.18
Grade math A	7.7	8.1	7.4	7.9
Grade math B	6.3	6.0	6.4	6.4
No math				
Share	0.02	0	0.01	0.04
Subjective probability to pass first year if assigned to ...				
- High reward	0.71	0.74	0.69	0.71
- Low reward	0.63	0.65	0.60	0.63
- Control	0.55	0.57	0.53	0.55
N	249	83	84	82

Note: Higher parental education includes university education and higher vocational education, lower parental education includes all other types of education.

tigate the statistical power we have performed some simulation studies (with a significance level equal to 0.05). Assuming a passing rate for students in the control group of 0.20 (which roughly coincides with historical passing rates), we investigate three possible situations.

First, assume that the rewards affect the passing rates of high reward group twice as much as the passing of the low reward group. The statistical power of a chi-square test for the null hypothesis that all groups have similar passing rates against the alternative hypothesis that each group has a different passing rate is about 0.5 if the passing rate of the high reward group becomes 0.31 (the passing rate in the low reward group is then 0.26). The power is about 0.8 if the passing rates in the high and the low reward group become 0.37 and 0.29 respectively.

Second, consider the case where the reward affects the passing rates, but the size of the reward is irrelevant, i.e. the passing rates of both reward groups are similar. To obtain a statistical power of 0.5 of a chi-square test for the null hypothesis that all groups have the same passing rates against the alternative hypothesis that the passing rates in the reward groups differ for passing rates in the control group, the passing rates in the reward groups should increase to 0.30. For a statistical power of 0.8, the passing rates for the reward groups should increase to 0.35.

Finally, if the reward only affects the students in the high reward group, for a statistical power of 0.5, the passing rates in the high reward group should increase to 0.32 and for a power of 0.8 this passing rate should become 0.37. From these calculations it is clear that the increase in passing rates necessary to obtain some reasonable statistical power is well within the 0.22-increase in passing rates that was found in the earlier study among econometrics students.

Table 2: Outcomes by group

	All (1)	High (2)	Low (3)	Control (4)
Passing rate	0.21	0.23	0.20	0.20
	(0.026)	(0.045)	(0.045)	(0.045)
Credit points	23.0	23.3	22.7	23.1
	(0.95)	(1.65)	(1.65)	(1.67)

Note: Mean values, with their standard errors in parentheses.

4. Results

4.1 Achievement

In Table 2 we report the first year passing rates and the average numbers of credit points for the full sample and for every treatment group. The passing rate of the high reward group is 0.03 higher than that of the low reward group and of the control group. However, the differences between the groups are not statistically significant. The p-value of a chi-square test for the null hypothesis that students in all three groups have identical passing rates against the alternative hypothesis that passing rates differ between groups equals 0.83. The p-value of this test against the alternative hypothesis that only individuals in the high reward group have different passing rates equals 0.54.[9]

Students in the control group collected on average 0.4 credit points more than students in the low reward group and 0.2 points less than students in the high reward group. Figure 1 shows the frequency distributions of the numbers of credit points for each of the three groups. The shapes of the three distributions are very similar. All three are bi-modal with the highest peak near 42 points and a second peak at the other extreme of 0 points. The p-value of a chi-square test for equality

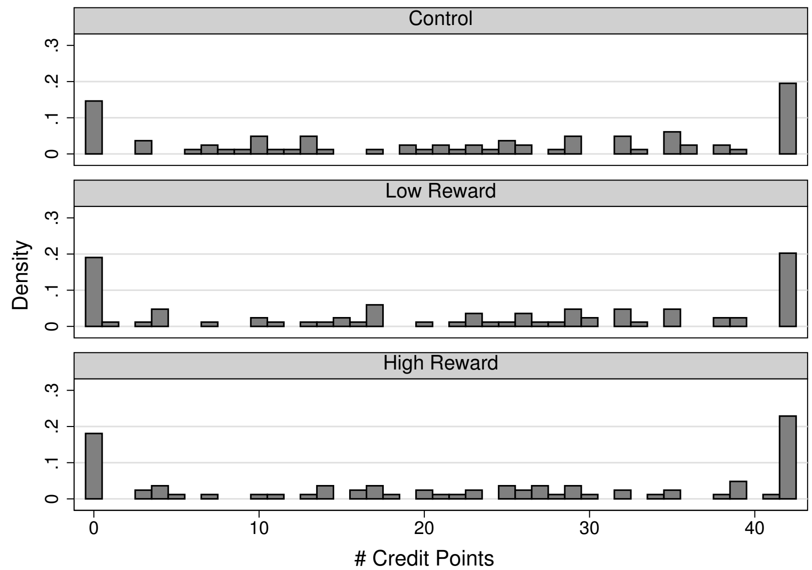

Figure 1: Distribution of total credit points collected by group

of the groups equals 0.97, indicating that there are no differences between the number of credit points collected by the different groups.

Finally, Figure 2 shows the average numbers of credit points per group accumulated over the course of the study year. The patterns are basically identical. The first key finding of this paper is therefore that the average treatment effect of the financial rewards on students' achievement is zero. This contradicts the finding of the earlier study among econometrics students.[10]

4.2 Effort and time allocation

The effect of rewards on achievement is a reduced form effect. It does not disentangle the effects of rewards on effort and subsequently of effort on achievement. To examine whether the zero effect of rewards on achievement is the result of the rewards having no impact on effort or the result of extra effort having no impact on achievement, we collected information about students' effort levels.

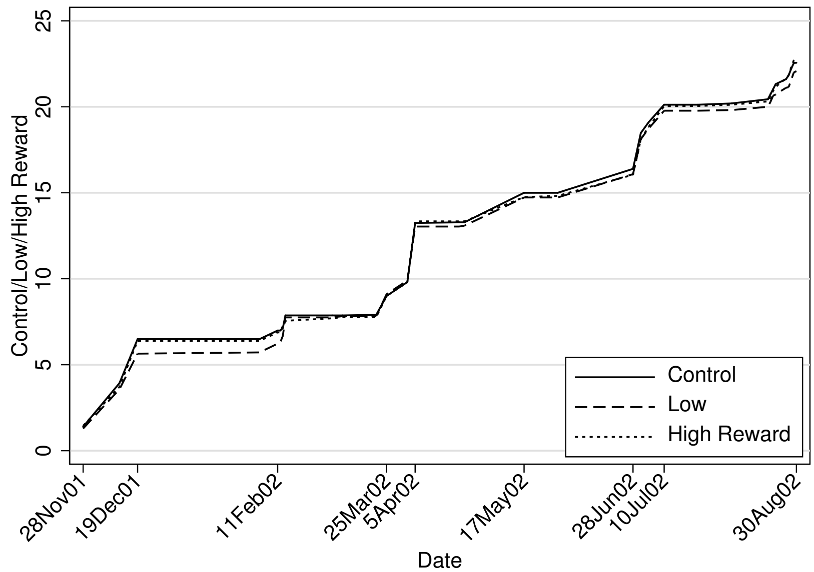

Figure 2: Average numbers of credit points collected during the academic year by group

The post-experiment questionnaire included the following questions:

- "How many hours per week did you on average spend on your study in economics and business during each of the three trimesters of the past academic year (2001/2002)? (We want to know the total average time spent on your study, this means including following and preparing lectures and courses and preparing exams.)"

- "How many hours did you spend in total on preparing make-up exams held in August? (Here we want to know the total number of hours, not the average per week.)"

Information about study time is provided in the first block of Table 3.

In all three groups, average study time is around 22.5 hours per week during the first trimester and decreases to around 18 during the second trimester and

Table 3: Time allocation by group

	All	High	Low	Control
Time spent on study				
First trimester (per week)	22.5	21.1	22.9	23.7
	(0.8)	(1.4)	(1.3)	(1.4)
Second trimester (per week)	18.2	18.2	17.7	18.9
	(0.8)	(1.4)	(1.3)	(1.4)
Third trimester (per week)	16.6	16.1	16.9	16.8
	(0.8)	(1.4)	(1.4)	(1.4)
Summer period (total)	27.4	30.5	22.5	29.5
	(2.3)	(4.1)	(3.9)	(4.1)
Effort increased as result of reward		0.37	0.21	
Paid job				
Share	0.80	0.81	0.83	0.76
	(0.03)	(0.05)	(0.04)	(0.05)
Hours worked (per week)	12.1	11.5	12.4	12.5
	(0.5)	(0.9)	(0.9)	(1.0)
Wage rate (in €)	7.60	7.22	7.88	7.69
	(0.17)	(0.30)	(0.29)	(0.31)
Member student association	0.26	0.29	0.26	0.23
	(0.03)	(0.05)	(0.05)	(0.05)
Living at parents' house	0.52	0.49	0.59	0.48
	(0.03)	(0.06)	(0.06)	(0.06)

Note: Sample means. Standard errors in parentheses.

16 to 17 during the third trimester. Students spend on average around 27 hours to prepare their make-up exams during the summer. Quite a few students report that they do not spend time at all on their study, which influences the averages for the second and third trimesters and for the summer period. These are the students who dropped out and for the summer period also students who did no make-up exams.[11] Average time spent on the study is very similar across groups, and only for the summer period average study time is highest for the high reward group. Differences across groups are not substantial nor statistically significant.

We are aware that the questions about study time measure actual effort only imperfectly. The responses are subjective and retrospective, and only measure time input and not the effective input per hour. While biases due to this may cancel out in across group comparisons, it is desirable to have additional information about study effort. The questionnaire therefore also included items concerning perceived increase in effort due to the reward, time spent on paid work, whether respondents joined a student association and whether they lived with their parents. Results are also reported in Table 3.

Thirty-seven percent of the participants who were actually assigned to the high reward group respond ex-post that the reward increased their study effort. In the low reward group this percentage equals 21. These two percentages are significantly different from each other. Hence, in the students' perception the rewards did not only affect study effort but also the size of the reward seems to matter.

Approximately 80 percent of the students combine studying with work, and those who work spend around 12 hours per week on this activity and earn on average €7.50 per hour. Here, we see no differences between the reward and control groups with the exception that students in the high reward group tend to earn somewhat lower wage rates than students in the other two groups. Finally, the last two rows of Table 3 reveal that the rewards did not withheld students from joining a student association or from moving out of their parents' house.

To summarize, the results on the effects of the financial rewards on effort levels are somewhat mixed. On the one hand we find no differences between groups in reported study time and other time allocation variables. It seems that many students prefer to combine work and study above devoting more time to their study. The rewards apparently do not change this preference and students do not shift time from work or leisure to their study. This result is consistent with the first finding that rewards do not affect achievement.[12] On the other hand, students' own perceptions point to increased effort by students in the reward groups, with a larger increase for higher rewards. Of course, these perceptions may be wrong and the answers may just reflect some socially desirable response. But it may also be the case that these responses reflect some genuine differences in exerted effort (per hour). If the latter is true, we must conclude that the rise is insufficient to increase the average student's achievement.

4.3 Heterogeneous treatment effects

So far we implicitly assumed that rewards affect all students identically. However, there are good reasons to expect that some students will be more responsive to a reward than others because of heterogeneity in the marginal cost of effort or heterogeneity in returns. Two student characteristics that seem particularly relevant in this respect are social background and ability.

Students from a poor social background may face credit constraints due to which they cannot afford a reduction of the time they work for pay. Consequently, these students are less likely to respond to the rewards than students from with more favorable social backgrounds. Students of high ability collect more credit points when there is no reward than low ability students and therefore have to bridge a smaller gap when a reward is promised. Moreover, high ability students earn more extra credits points with a given increase of their effort than low ability students. Consequently, high ability students are more likely to respond to the rewards than low ability students.

Let us first investigate to what extent the effect of the rewards depends on social background. We use the level of fathers' education as a measure of social background and split the sample in 133 students whose fathers completed at least higher education, and 116 students whose fathers do not have a higher education degree. Columns (2) and (3) in Table 4 show the passing rates and the number of credit points collected for both subsamples conditional on treatment status.

Of those whose fathers are higher educated, the students in the high reward group have a higher passing rate and they also collect more credit points than the students in the low reward group and the control group. In this subsample students in the high reward group thus tend to perform somewhat better than students in the other groups. Within the subsample of students whose fathers completed less than higher education the students in the high reward group do not have higher passing rates, nor do they collect more credit points than students in both other groups.

Next consider student ability. As an indicator for ability we use the secondary school math grades of the students. We split the sample in two: students with good math skills and students with poor math skills. We consider a student to have good math skills if either his secondary school grade for math A was 8 or higher or the grade for math B was 6 or higher. 107 students enter the high math skill group and 142 students the low math grade group. The passing rates and

Table 4: Outcomes by group and students' background characteristics

		Father's education		Math grade	
	All	High	Low	High	Low
		(2)	(3)	(4)	(5)
Passing rate					
- High reward	0.23	0.27	0.18	0.45	0.08
	(0.05)	(0.06)	(0.07)	(0.09)	(0.04)
- Low reward	0.20	0.18	0.23	0.39	0.09
	(0.04)	(0.06)	(0.06)	(0.09)	(0.04)
- Control	0.20	0.20	0.18	0.33	0.09
	(0.04)	(0.06)	(0.07)	(0.08)	(0.04)
Credit points					
- High reward	23.3	25.2	20.9	32.8	17.0
	(1.7)	(2.3)	(2.4)	(2.2)	(2.1)
- Low reward	22.7	20.1	25.6	29.5	18.8
	(1.7)	(2.3)	(2.3)	(2.3)	(2.0)
- Control	23.1	21.6	24.7	28.1	19.1
	(1.7)	(2.3)	(2.4)	(2.1)	(2.2)

Note: Sample means with their standard errors in parentheses.

the average number of credit points collected for both subsamples differentiated by reward and control groups are presented in columns (4) and (5) of Table 4.

Of those with high math skills, the students in the high reward group have higher passing rates and collect more credit point than those in the low reward group, who in turn perform better than those in the control group. Within the low math skills group differences between reward groups and control group are not very pronounced.

Due to the rather small sizes of the various sub-samples, the reward effects within the high social background and high ability sub-samples are not very precisely measured and lack statistical significance. Nevertheless, the patterns nicely concur with our expectations. Especially the effects of rewards for students with high math skills seem substantial and reveal a monotonic pattern. Without a reward, these students earn on average 28 credit points, and every €227 reward increases their average number of credit points by 1.5.

Because high ability students and students with higher social background have higher achievement if a (higher) reward is promised, we also expect them devote more time on their study. The results in of Table 5 lend no support for that. The first two blocks present information on study time broken down by sub-samples and treatment groups. For conciseness we report the average weekly study time during the three trimesters. Study time during the summer period is reported separately. Within each of the four sub-samples we observe virtually no differences in average weekly study time between reward and control groups. There are some differences in the amounts of study time during the summer especially for students with a lower parental education background. None of these differences is, however, statistically significant.

The bottom part of Table 5 gives - again by sub-sample and reward size -

Table 5: Study time and effort by treatment and students' background

		Father's education		Math grade	
	All	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)
Average weekly study time trimesters 1-3					
- High reward	18.4	18.0	19.0	19.0	18.1
	(1.2)	Effort	(1.6)	(1.7)	(1.6)
- Low reward	19.2	18.7	19.7	19.7	18.9
	(1.1)	(1.7)	(1.5)	(1.7)	(1.6)
- Control	19.8	18.3	21.4	19.6	19.9
	(1.2)	(1.8)	(1.6)	(1.6)	(1.7)
Study time during summer period					
- High reward	30.5	25.3	37.4	22.0	36.2
	(4.1)	(4.9)	(6.9)	(4.6)	(6.0)
- Low reward	22.5	20.0	25.1	22.3	22.6
	(3.9)	(5.0)	(6.2)	(4.6)	(5.7)
- Control	29.5	29.8	29.1	26.1	32.0
	(4.1)	(5.1)	(6.7)	(4.6)	(6.3)
Effort increased as result of reward					
- High reward	0.37	0.43	0.29	0.48	0.30
	(0.05)	(0.06)	(0.08)	(0.09)	(0.06)
- Low reward	0.21	0.12	0.30	0.32	0.14
	(0.05)	(0.07)	(0.07)	(0.09)	(0.06)

Note: Sample means with standard errors in parentheses.

the shares of students who respond that the rewards affected their study effort. Within the sub-samples of students with higher educated fathers, students with high math skills and students with low math skills, the fractions of students who say that the reward did have an impact is substantially larger in the high reward group than in the low reward group. These results provide an explanation for the heterogeneous treatment effects reported in Table 4. High ability students and students with higher social backgrounds who can earn a higher reward have higher achievement because they put more effort (per hour) into their study. Low ability students who are promised the high reward put more effort into their study than they would otherwise have done, but apparently this does not boost their achievement.

The insights gained from the results for various sub-samples readily explain why we find a negligible average treatment effect. Students who can afford to forgo earnings from part-time work and students for whom the reward requirement is feasible increase their (perceived) effort and achieve more. Students for whom the reward requirements are too demanding and students who cannot afford to reduce the time they work for pay are not responsive to the rewards.

For high ability students and students from a high social background the requirements are obviously feasible. They can afford to forgo earnings from paid work and increase their (perceived) effort and achieve more. These insights readily explain why we find indications for positive treatment effects for some subgroups but a negligible average treatment effect.

4.4 Threats to validity

While a randomized experiment is often considered the gold standard in research on treatment evaluation, it is not without threats to validity of the outcomes.

Table 6: Incidence and size of supplementary rewards

	Incidence rate	Mean reward size
High reward	0.104	€770
Low reward	0.025	€750
Control	0.053	€625

Heckman et al. (1999) and Philipson (2000) have drawn attention to the importance of general equilibrium effects and external treatment effects or spillover effects. In the context of our experiment at least three confounding factors may play a role. First there may be treatment substitution bias. Parents may promise a reward or may supplement the reward if the students are assigned to the control or low reward group. In this case all participants could be confronted with essentially the same treatment and we would most likely find no difference between the original three groups. To investigate whether such responses actually took place, we included in the post-experiment questionnaire a question whether someone else (for instance parents) promised a reward for passing all first year exams. Table 6 reports for each group the shares of students responding affirmative to this question along with the mean values of the size of these supplementary rewards. The table shows that supplementary rewards are fairly uncommon, and that incidence rate and size of such rewards are higher among the high reward group than among the low reward group and the control group. Therefore we expect supplementary rewards to have no impact on our findings.

A second possible confounding factor is that teachers may grade exams differently for students in the reward groups than for students in the control groups. Although teachers are in principle unaware of the treatment status of their students, students could communicate their status in the hope that teachers will grade their exams more favorably. This seems unlikely for two reasons. First, students from the control group could also claim that they belong to a reward group if this implies that their exam will be graded more favorably.[13] A second and more important reason is that during the first academic year most exams are multiple-choice tests. Such tests give teachers little leeway to manipulate grades of particular students.

A final possible confounding factor is that if the rewards induce students in the reward groups to work harder, that this could spill over to their peers in the control group. During the design phase of the experiment we considered the possibility of a two-stage randomization scheme as proposed by Philipson (2000). First year students in economics and business at the University of Amsterdam are placed into different classes. Students placed in the same class are supposed to follow the first year program together. That means: having the same weekly schedule, having the same teachers and having the same peers, but the exams are always the same for all students. In the academic year 2001/2002, there were 9 of such classes. This would allow to first assign different intensities of treatment to different classes, and then within classes assign students to reward and control groups.

This two-stage randomization could be undermined by the fact that assignment into classes is not random, but depends on math achievement in secondary school and the order of application for the economics and business study. The order of application is likely to be related with students' motivation. The design could then lead in practice to unbalanced reward and control groups where treatment assignment is not orthogonal to unobserved students' characteristics. Furthermore, students often do not stay in the class of their original assignment, thereby generating a similar confounding effect. Finally, a two-stage design would make it more difficult if not impossible to convince participants that the odds to be assigned to one of the reward groups were equal for all.

We consider it unlikely that spillover effects influenced our findings. The overall passing rate of the students in our experiment is identical to the passing rates of previous cohorts. Information about student effort from previous cohorts is in line with student effort among the students that participated in the experiment. There is also no change in the composition of the student population in terms of secondary school grades for mathematics.

5. Conclusion

This paper reports about a randomized social experiment that investigated the effects of financial incentives on undergraduate students' achievement. The target population consists of first year economics and business students at the University of Amsterdam. The students, who were randomized in the reward groups, were promised a reward upon passing all first year exams before the start of their second academic year. In the high reward group the reward was €681 and in the low reward group the reward was €227. Students in the control group could not earn a reward.

The results of the experiment point to a negligible average treatment effect. Measured by passing rates and collected credit points, students in the reward groups have the same achievement levels as students in the control group. If we examine the results for different sub-samples, we find, however, some evidence for heterogeneous treatment effects. Students possessing good math skills and students with more highly educated fathers do respond to the rewards. Especially for students with high math skills, the 0.06 and 0.12 increases in the passing rate caused by the low and high reward respectively, are substantial.

The heterogeneous responses also help explain why we find an average treatment effect equal to zero. Apparently, only students for whom the reward requirement is feasible respond to the incentives. This explanation is similar to that of Angrist and Lavy (2002), who find no (their first experiment) or only small (their second experiment) average treatment effects. The students in their sample come from a disadvantaged group with very low initial (no reward) passing rates.

In a post-experiment questionnaire, we also collected information on students' time allocation and study effort. This additional information gives mixed insights. On the one hand, students' reported study time is not affected by the rewards. On the other hand, students in the reward groups claim that they worked harder as a result of the rewards. The pattern of this latter indicator of effort across sub-samples and treatment groups is consistent with observed differences in achievement.

Our experiment was conducted at a Dutch university with a cohort of economics and business students. Like with any other social experiment, the extent to which the results can be generalized to other populations (countries, universities, groups of students, etc.) or treatments is unclear. Our finding that financial rewards do not boost average students' achievements does therefore not imply that financial rewards will never improve students' achievements. In our interpretation, the reward requirements in the experiment were too demanding for the average economics and business student in relation to the size of the rewards. The indicative finding of heterogeneous treatment effects suggests to us that the effectiveness of financial rewards for students depends crucially on the feasibility of the requirements for collecting such rewards.

References

Angrist, J. D., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Columbia: Evidence from a randomized natural experiment. *American Economic Review*, 92(5):1535–1558.

Angrist, J. D. and Lavy, V. (2002). The effect of high school matriculation awards: Evidence from randomized trials. Working Paper Series 9389, NBER.

Dearden, L., Emmerson, C., Frayne, C., Goodman, A., Ichimura, H., and Meghir, C. (2001). Education maintenance allowances: The first year - a quantitative evaluation. Research Report RR257, DfES.

Dearden, L., Emmerson, C., Frayne, C., and Meghir, C. (2002). Education maintenance allowances: The first two years - a quantitative evaluation. Research Report RR352, DfES.

Heckman, J. J., Lochner, L., and Taber, C. (1999). General equilibrium cost benefit analysis of education and tax policies. Working Paper Series 6881, NBER.

Hilkhuysen, P. (2000). Premie studieresultaat; experiment in de propedeuse AEO: cohort 1999/2000. Jaarverslag onderwijsinstituut 1999-2000, bijlage 7, Faculteit der Economische Wetenschappen en Econometrie, Universiteit van Amsterdam.

Oosterbeek, H., Groot, W., and Hartog, J. (1992). An empirical analysis of university choice and earnings. *De Economist*, 140:293–309.

Philipson, T. J. (2000). External treatment effects and program implementation bias. Technical Working Paper Series T0250, NBER.

Notes

[1]In the Netherlands, econometrics is a separate undergraduate education from economics and business.

[2]The first draft of this study circulated only after our experiment started.

[3]Two other programs that provide financial incentives for achievement are the Education Maintenance Allowance (EMA) in Britain and the Programa de Ampliacion de Cobertura de la Educacion Secundaria (PACES) in Columbia. EMA gives low-income families a payment for enrollment and achievement. Assignment to treatment is, however, not random. Dearden et al. (2001, 2002) describe the evaluation of this program. PACES is a program in which more than 125,000 Columbian pupils received vouchers which covered about half of the cost of private secondary school. Vouchers were only renewed for pupils who maintained satisfactory academic performance (Angrist et al., 2002).

[4]Students who successfully completed the first year in a higher vocational school can also enter university. These students are a small fraction of the total inflow into university.

[5]For a few studies students are admitted on the basis of a lottery when the number of applicants exceeds the number of available places. This is not the case for the economics and business studies.

[6]These are: Erasmus University Rotterdam, Free University Amsterdam, University of Amsterdam, University of Groningen, University of Maastricht and University of Tilburg.

[7] Table A1 in the appendix gives an overview of the first year courses and the

number of credit points (weeks) assigned to each course.

[8]Students can receive 1 credit point dispensation for part of the financial accounting course if they followed a specific course during secondary education.

[9]The post-experiment questionnaire also asked students whether they were still studying economics and business. One quarter of the participants respond that they dropped out; this share does not differ between reward and control groups.

[10]Due to the timing of the experiment one could argue that students' achievement on the December exams are pre-program outcomes and use these outcomes as such. Figure 2 makes clear that reward and control groups perform similar on both the December exams (pre-program outcomes) and all the subsequent exams (program outcomes).

[11]In the first trimester 3 respondents report zero study effort, in the second trimester this equals 33 and in the third trimester 39; 83 students spent zero hours on preparing for the August make-up exams, of which 22 students did not have to do any make-up exams. For the sample reporting positive numbers, the distribution of study time is bell-shaped.

[12]If the rewards would have increased students' study time then we would have been able to estimate the causal effect of study time on achievement. Since the rewards do not change study time we cannot estimate such an effect. Regressing the passing rate on study time we find that one hour study time extra per week is associated with a one percent higher passing rate. Adding controls for ability, social background and the subjective passing rate does not change the size of this correlation.

[13]It is unlikely that teachers demand of students to prove to which group they belong.

Table A1: Overview of the first year courses in the economics and business program

	Credit points
Trimester 1 (September-December)	
- Financial accounting	4
- Microeconomics	6
- Mathematics 1	3
- Information management A	1
Trimester 2 (January-March)	
- Macroeconomics	6
- Management accounting	3
- Orientation fiscal economics	1
- Mathematics 2	3
- Information management B	1
Trimester 3 (April-June)	
- Finance	4
- Marketing	3
- Organization	3
- Statistics	3
- Information management C	1